

FIGURE 1

|  |     |
|--|-----|
| ATGGGTTCCATGCGTCTATT   | 20  |
| <u>M G S M R L L</u>   |     |
| ----- prx9+ ----->   |     |
| AGTAGTGGCATTGTTGTGTGCATTTGCTATGCATGCAGGTTTTTCAGTCTCTTATGCTCA   | 80  |
| <u>V V A L L C A F A M H A G F S V S Y A Q</u>                 | 1   |
| signal sequence  |     |
| GCTTACTCCTACGTTCTACAGAGAAACATGTCCAAATCTGTTCCCTATTGTGTTTGGAGT   | 140 |
| L T P T F Y R E T C P N L F P I V F G V                        | 21  |
| ----- prx12+ ----->  |     |
| AATCTTCGATGCTTCTTTACCGATCCCCGAATCGGGGCCAGTCTCATGAGGCTTCATTT    | 200 |
| I F D A S F T D P R I <u>G A S L M R L H F</u>                 | 41  |
| active site  |     |
| I <-----   |     |
| TCATGATTGCTTTGTTCAAG GTTGTGATGGATCAGTTTTGCTGAACAACACTGATACAAT  | 260 |
| <u>H D C F V Q G C D G S V L L N N T D T I</u>                 | 61  |
| --prx10- --- ----- prx2+ ----->                                |     |
| AGAAAGCGAGCAAGATGCACTTCCAAATATCAACTCAATAAGAGGATTGGACGTTGTCAA   | 320 |
| E S E Q D A L P N I N S I R G L D V V N                        | 81  |
| TGACATCAAGACAGCGGTGGAAAATAGTTGTCCAGACACAGTTTCTTGTGCTGATATTCT   | 380 |
| D I K T A V E N S C P D T V S C A D I L                        | 101 |
| II   |     |
| TGCTATTGCAGCTGAAATAGCTTCTGTTCTG GGAGGAGGTCCAGGATGGCCAGTTCCATT  | 440 |
| A I A A E I A S V L G G G P G W P V P L                        | 121 |
| AGGAAGAAGGGACAGCTTAACAGCAAACCGAACCTTGCAAATCAAAACCTTCCAGCACC    | 500 |
| G R R D S L T A N R T L A N Q N L P A P                        | 141 |
| TTTCTTCAACCTCACTCAACTTAAAGCTTCCTTTGCTGTTCAAGGTCTCAACACCCTTGA   | 560 |
| F F N L T Q L K A S F A V Q G L N T L D                        | 161 |
| III  |     |
| TTTAGTTACACTCTCAG GTGGTCATACGTTTGGGAAGAGCTCGGTGCAGTACATTCATAAA | 620 |
| <u>L V T L S G G H T F</u> G R A R C S T F I N                 | 181 |
| heme-binding domain  |     |
| CCGATTATACAACCTTCAGCAACACTGGAAACCCTGATCCAACCTCTGAACACAACATACTT | 680 |
| R L Y N F S N T G N P D P T L N T T Y L                        | 201 |
| AGAAGTATTGCGTGCAAGATGCCCCCAGAATGCAACTGGGGATAACCTCACCAATTTGGA   | 740 |
| E V L R A R C P Q N A T G D N L T N L D                        | 221 |
| CCTGAGCACACCTGATCAATTTGACAACAGATACTACTCCAATCTTCTGCAGCTCAATGG   | 800 |
| L S T P D Q F D N R Y Y S N L L Q L N G                        | 241 |
| CTTACTTCAGAGTGACCAAGAACTTTTCTCCACTCCTGGTGCTGATACCATTTCCCATTTGT | 860 |

|   |      |
|---|------|
| L L Q S D Q E L F S T P G A D T I P I V                       | 261  |
| <----- prx6- -----  |      |
| CAATAGCTTCAGCAGTAACCAGAATACTTTCTTTTCCAACCTTTAGAGTTTCAATGATAAA | 920  |
| N S F S S N Q N T F F S N F R V S M I K                       | 281  |
| AATGGGTAATATTGGAGTGCTGACTGGGGATGAAGGAGAAATTCGCTTGCAATGTAATT   | 980  |
| M G N I G V L T G D E G E I R L Q C N F                       | 301  |
| TGTGAATGGAGACTCGTTTGGATTAGCTAGTGTGGCGTCCAAAGATGCTAAACAAAAGCT  | 1040 |
| V N G D S F G L A S V A S K D A K Q K L                       | 321  |
| TGTTGCTCAATCTAAATAAACCAATAATTAATGGGGATGTGCATGCTAGCTAGCATGTAA  | 1100 |
| V A Q S K *   | 326  |
| AGGCAAATTAGGTTGTAAACCTCTTTGCTAGCTATATTGAAATAAACCAAAGGAGTAGTG  | 1160 |
| TGCATGTCAATTCGATTTTGCCATGTACCTCTTGGAATATTATGTAATAATTATTTGAAT  | 1220 |
| CTCTTTAAGGTACTTAATTAATC (A) n                                 |      |

FIGURE 2

|      | 10   | 20 | 30 | 40 | 50 | 60 |
|------|--|----|----|----|----|----|
| 1    | GCATCATATCATAAACAATACGTACGTGATATTATCTAGTGTCTCTCAGTTTACTTTATG     |    |    |    |    |    |
| 61   | AGAAATTATTTTCTTTAAAAAAGTTAATTAATAAAAACATTTGCGATACCGTGAGTTA       |    |    |    |    |    |
| 121  | CAAGAAATCCGCCGAATTCATCTCTATAAATAAAAGGATCTATATGAGAGGTAAAATCAT     |    |    |    |    |    |
| 181  | ATTAAC TCAAAATGGGTTC CATGCGTCTATTAGTAGTGGCATTGTTGTGTGCATTTGCTA   |    |    |    |    |    |
| 241  | TGCATGCAGGTTTTTCAGTCTCTTATGCTCAGCTTACTCCTACGTTCTACAGAGAAACAT     |    |    |    |    |    |
| 301  | GTCCAAATCTGTTCCCTATTGTGTTTGGAGTAATCTTCGATGCTTCTTTCCACCGATCCCC    |    |    |    |    |    |
| 361  | GAATCGGGGCCAGTCTCATGAGGCTTCATTTTCATGATTGCTTTGTTCAAGTACGTACTT     |    |    |    |    |    |
| 421  | TTTTTTTTCTTCCAAAATGCCCTGCATATTTAACAAGATTGCTTTGTTCCACCTAGAAAA     |    |    |    |    |    |
| 481  | ATGTGTTTTTTTTCAACGATCTTACGTACGTTTGTGTTTGGTTTGA AAAATAAATCAGAAAGA |    |    |    |    |    |
| 541  | GATCAAGAAAATAGCTAGAAAGAAAGCAACGTTTTTTTTAAAAGGTATTTAGTGTGAGAAA    |    |    |    |    |    |
| 601  | AATATTA AAACTGAAGAGAAAGAAATTAATAAGCTTTTCTTGAATGATATTTACATGTC     |    |    |    |    |    |
| 661  | TTATTAAC TTAAGTCACCTTTTTTCTTTAAGTTGTGCTTGAAGAAAAAAGATGTCTTTC     |    |    |    |    |    |
| 721  | AGTTTAGTTTTGATTAATGCTAATTATTTTTTAATTAATTAATACTATATATCTA          |    |    |    |    |    |
| 781  | TTTACCATATTAATTATTACTATATTTTCATGATGACAACAGACAAGTATTCTAAAGAGGT    |    |    |    |    |    |
| 841  | ATCGGTAGATGATTAATTTTTTTATAAAAAAATCTTTTGCGTGTATAGATATTCTTTTAT     |    |    |    |    |    |
| 901  | AATTGGTGCAGAACTTGTAATGCTAATTGCAATTAATCTTACATTGATTAAC TAATAGC     |    |    |    |    |    |
| 961  | TATAATCAATATTTAGGTTAGGTATAGGAGACAAATCAAGTGATCTGAACAAATTAAGTT     |    |    |    |    |    |
| 1021 | GTTATATTTGCATTGTGACAGGGTTGTGATGGATCAGTTTTGCTGAACAACACTGATACA     |    |    |    |    |    |
| 1081 | ATAGAAAGCGAGCAAGATGCACTTCCAAATATCAACTCAATAAGAGGATTGGACGTTGTC     |    |    |    |    |    |
| 1141 | AATGACATCAAGACAGCGGTGGAAAATAGTTGTCCAGACACAGTTTCTTGTGCTGATATT     |    |    |    |    |    |
| 1201 | CTTGCTATTGCAGCTGAAATAGCTTCTGTTCTGGTAATTAATAACTCCTAATTAATTCCC     |    |    |    |    |    |
| 1261 | AACCATTAAAAAGTTGCATGATTGGATTCAAAATTCTATGGTATTGGGGTTCTGATATAA     |    |    |    |    |    |
| 1321 | ATTTGTAATTA AATTGCACTAAAAAAAATTATCATATACTTTTAATAAAAAAAATTTATC    |    |    |    |    |    |
| 1381 | TAATTTAATTTATTATTA AAAACTATTTTTAAAATTCAATCCTAACTCTTTTTTAATCGGA   |    |    |    |    |    |
| 1441 | GCATGTAAGCTGGCACCCACCGTATATCGTTGGAAGATGCTATAAAACCATTTAATTAAT     |    |    |    |    |    |
| 1501 | GGATGGAATCAGTCAAAACATTTAATTCAAAATACTCTTAATTGTGATTAGTAATCATGT     |    |    |    |    |    |
| 1561 | TCGGGCAAGTTACGTTGTGTATAATTAATTTGACTTAATCAGATAAAAAAACAATGGAC      |    |    |    |    |    |
| 1621 | GCAAGCCGGTTGGTATAGATATCACTGGCCTGTAGAATATGTGGTTTTTTCAGTTTTAAAT    |    |    |    |    |    |
| 1681 | AAAAGCTAGCTACTATATTATATTTAGTCTTTTTTTTTCTTAAACCCATTTAACGTGATT     |    |    |    |    |    |
| 1741 | TATTGACTGTGAAACATGTTTCCACACACAGGCTTAGAAACTCCTCGCAACTAACATCTC     |    |    |    |    |    |
| 1801 | CAAAATTTGACTATTTATTTATGAAGATAATTCATCTATGATGTTCAACTCTATTATATA     |    |    |    |    |    |
| 1861 | TATGTATCATCGCAGTATTAAGAATTATAATAGTCAAATATAGAAGTATATCGGGTAAAT     |    |    |    |    |    |
| 1921 | GTAGTTGCATGTGCGACCTGTTTCGTGTAAAATGCTTATTCTATATAGCTTTTTTTTATTG    |    |    |    |    |    |
| 1981 | GAAAATAACGATGAACTAAAAACGAAAGGGTATCATATAGTTTGACTTTTATGTTAGAGA     |    |    |    |    |    |
| 2041 | GAGACATCTTAATTTGGTCATATGTTAAATAATTAATTACAATGCATACACAAATATTTA     |    |    |    |    |    |
| 2101 | TGCCATATCTAAAAAATGATAAAATATCATAGGTATACTCAACTATATGATATCCCCATA     |    |    |    |    |    |
| 2161 | ACAGAAATTGTACTTTTCTTCAGGCAATGAACTTAACATTTCTGTTTGCTAAAAACAAAC     |    |    |    |    |    |
| 2221 | ATCCACTTAAAGTGGTTCAACATATTTATGTAATAATTTACAGGGAGGAGGTCCAGGATG     |    |    |    |    |    |
| 2281 | GCCAGTTCCATTAGGAAGAAGGGACAGCTTAACAGCAAACCGAACCTTGCAAATCAAAA      |    |    |    |    |    |
| 2341 | CCTTCCAGCACCTTTCTTCAACCTCACTCAACTTAAAGCTTCCTTTGCTGTTCAAGGTCT     |    |    |    |    |    |
| 2401 | CAACACCCTTGATTTAGTTACACTCTCAGGTATACATAATCAATTTTTTATTTGCTATTA     |    |    |    |    |    |
| 2461 | GCTAGCAATAAAAAGTCTCTGATACAGACATATTTAGATAAATTAATTTCTCCATAAACA     |    |    |    |    |    |
| 2521 | TTTATAATAAAATTATCAATTTATGTACTTAAAAATTATGGATTGAAGCTCTTTTCATCC     |    |    |    |    |    |
| 2581 | AACTTTTACTAAAGTTAAGGTGCATATAATATAAAATAAACTATCTCTTGTCTTATAA       |    |    |    |    |    |
| 2641 | AAAGATTGAAGATAAGTTAAAGTCTACTTATAAATCATTAATATATGTATAGGTGGTCAT     |    |    |    |    |    |
| 2701 | ACGTTTGGAAGAGCTCGGTGCAGTACATTCATAAACCGATTATACAACCTTCAGCAACACT    |    |    |    |    |    |
| 2761 | GGAAACCCTGATCCAACCTCTGAACACAACATACTTAGAAGTATTGCGTGCAAGATGCCCC    |    |    |    |    |    |
| 2821 | CAGAATGCAACTGGGGATAACCTCACCAATTTGGACCTGAGCACACCTGATCAATTTGAC     |    |    |    |    |    |
| 2881 | AACAGATACTACTCCAATCTTCTGCAGCTCAATGGCTTACTTCAGAGTGACCAAGA ACTT    |    |    |    |    |    |
| 2941 | TTCTCCACTCCTGGTGCTGATACCATTCCCATTTGTCAATAGCTTCAGCAGTAACCAGAAT    |    |    |    |    |    |
| 3001 | ACTTTCTTTTCCAACCTTAGAGTTTCAATGATAAAAATGGGTAATATTGGAGTGCTGACT     |    |    |    |    |    |

3061 GGGGATGAAGGAGAAATTCGCTTGCAATGTAATTTTGTGAATGGAGACTCGTTTGGATTA  
3121 GCTAGTGTGGCGTCCAAAGATGCTAAACAAAAGCTTGTTGCTCAATCTAAATAAACCAAT  
3181 AATTAATGGGGATGTGCATGCTAGCTAGCATGTAAAGGCAAATTAGGTTGTAAACCTCTT  
3241 TGCTAGCTATATTGAAATAAACCAAAGGAGTAGTGTGCATGTCAATTCGATTTTGCCATG  
3301 TACCTCTTGGAATATTATGTAATAATTATTGTAATCTCTTTAAGGTACTTAATTAATCA

FIGURE 3A

|        |   |     |
|--------|---|-----|
| L78163 | -----ATGGGTTCATGCGT-CTATTAGTAGTGGCATTGTTG                     | 36  |
| U41657 | -----   | 0   |
| X90693 | G----GCAAA-CAATGAACTCCCTTCGTGCTGTAGCAATAG-CTTTGTGC            | 44  |
| X90694 | GCTCTTCAAAACAATGAACTCC-----TTAGCAACTT-CTATGTGG                | 40  |
| L36156 | -----CTCC-----TTAGCAACTT-CTATGTGG                             | 22  |
| X90692 | -----AATGCTTGGT-----CTAAGTGCAACAGCTTTTTGCTGTATGG              | 38  |
| L78163 | TGT-----GCATTT-GCTATGCATGCAGGTTTTTCAGT---CTCTTATGC            | 77  |
| U41657 | -----   | 0   |
| X90693 | TGTATTGTG-----GTTGTGCTTGGAGGGTTACCCTTCTCTTCAAATGC             | 88  |
| X90694 | TGTGTTGTGCTTTTAGTTGTGCTTGGAGGACTACCCTTTTCCTCAGATGC            | 90  |
| L36156 | TGTGTTGTGCTTTTAGTTGTGCTTGGAGGACTACCCTTTTCCTCAGATGC            | 72  |
| X90692 | TGT-TTGTGCTAAT-----TGGAGGAGTACCCTTTT---CAAATGC                | 75  |
| L78163 | TCAGCTTACTCCTACGTTCTACAGAGAAACATGTCCAAATCTGTTCCCTA            | 127 |
| U41657 | -----   | 0   |
| X90693 | GCAACTTGATCCATCCTTTTACAGGAACACTTGTCCAAATGTTAGTTCCA            | 138 |
| X90694 | ACAACCTAGTCCCACCTTTTACAGCAAAACGTGTCCAACTGTTAGTTCCA            | 140 |
| L36156 | ACAACCTAGTCCCACCTTTTACAGCAAAACGTGTCCAACTGTTAGTTCCA            | 122 |
| X90692 | ACAACCTAGATCCTTCATTTTACAACAGTACATGTTCTAATCTTGATTCAA           | 125 |
| L78163 | TTGTGTTTGGAGTAATCTTCGATGCTTCTTTCACCGATCCCCGAATCGGG            | 177 |
| U41657 | -----   | 0   |
| X90693 | TTGTTCGTGAAGTCATAAGGAGTGTTTCTAAGAAAGATCCTCGTATGCTT            | 188 |
| X90694 | TTGTTAGCAATGTCTTAACAAACGTTTCTAAGACAGATCCTCGCATGCTT            | 190 |
| L36156 | TTGTTAGCAATGTCTTAACAAACGTTTCTAAGACAGATCCTCGCATGCTT            | 172 |
| X90692 | TCGTACGTGGTGTGCTCACAAATGTTTACAATCTGATCCCAGAATGCTT             | 175 |
| L78163 | GCCAGTCTCATGAGGCTTCATTTTCATGATTGCTTTGTTCAAGGTGTGA             | 227 |
| U41657 | -----TTTCATGATTGCTTTGTTCAAGGTGTGA                             | 29  |
| X90693 | GCTAGTCTTGTGTCAGGCTTCACTTTTCATGACTGTTTTGTTCAAGGTGTGA          | 238 |
| X90694 | GCTAGTCTCGTCAGGCTTCACTTTTCATGACTGTTTTGTTCTGGGATGTGA           | 240 |
| L36156 | GCTAGTCTCGTCAGGCTTCACTTTTCATGACTGTTTTGTTCTGGGATGTGA           | 222 |
| X90692 | GGTAGTCTCATCAGGCTACATTTTCATGACTGTTTTGTTCAAGGTGCGA             | 225 |
|        | ***** ** ***** . . . ** ** *                                  |     |
| L78163 | TGGATCAGTTTTGCTGAACAACACTGATACAATAGAAAGCGAGCAAGATG            | 277 |
| U41657 | TGGATCAGTTTTACTGAACAACACTGATACAATAGAAAGCGAGCAAGATG            | 79  |
| X90693 | TGCATCAGTTTTACTAAACAAACACTGATACCGTTGTGAGTGAACAAGATG           | 288 |
| X90694 | TGCCTCAGTTTTGCTGAACAATACTGCTACAATCGTAAGCGAACAACAAG            | 290 |
| L36156 | TGCCTCAGTTTTGCTGAACAATACTGCTACAATCGTAAGCGAACAACAAG            | 272 |
| X90692 | TGCCTCGATTTTTGCTGAACGATACGGCTACAATAGTGAGCGAGCAAAGTG           | 275 |
|        | ** ** . . ***** . ** . ***** * ** . * ** . * ** . ** . ** . * |     |
| L78163 | CACTTCCAAATATCAACTCAATAAGAGGATTGGACGTTGTCAATGACATC            | 327 |
| U41657 | CACTTCCAAATATCAACTCAATAAGAGGATTGGACGTTGTCAATGACATC            | 129 |
| X90693 | CTTTTCCAAACAGAACTCATTAAGAGGTTTGGATGTTGTGAATCAAATC             | 338 |
| X90694 | CTTTTCCAAATAACAACCTCTCTAAGAGGTTTGGATGTTGTGAATCAGATC           | 340 |
| L36156 | CTTTTCCAAATAACAACCTCTCTAAGGGGTTTGGATGTTGTGAATCAGATC           | 322 |
| X90692 | CACCACCAATAACAACCTCCATAAGAGGTTTGGATGTGATAAACCAGATC            | 325 |
|        | * . . ***** * . ***** ***** . ** . * ** * ** *                |     |
| L78163 | AAGACAGCGGTGGAAAATAGTTGTCCAGACACAGTTTCTTGTGCTGATAT            | 377 |

|        |  |     |
|--------|--|-----|
| U41657 | AAGACAGCGGTGGAAAATAGTTGTC CAGACACAGTTTCTTG TGCTGATAT   | 179 |
| X90693 | AAACAGCTGTGGAAAAGGCTTGT CCTAACACAGTTTCTTG TGCTGATAT  | 388 |
| X90694 | AAACTGGCTGTAGAAGTGCCTTGT CCTAACACAGTTTCTTG TGCTGATAT   | 390 |
| L36156 | AAACTGCTGTAGAAGTGCCTTGT CCTAACACAGTTTCTTG TGCTGATAT  | 372 |
| X90692 | AAACAGCGGTGGAAAATGCTTGT CCTAACACAGTTTCTTG TGCTGATAT<br>** . **.*.*.*.*.. *****.*****                       | 375 |
|        |  |     |
| L78163 | TCTTGCTATTGCAGCTGAAATAGCTTCTGTT- CTGGGAGGAGGTCCAGGA  | 426 |
| U41657 | TCTTGCTATTGCAGCTGAAATAGCTTCTGTTGCTGGGAGGAGGTC- AGGA  | 228 |
| X90693 | TCTTGCTCTTTCTGCTGAATTATCATCTACA- CTGGCAGATGGTCCTGAC  | 437 |
| X90694 | TCTTGCACTTGCTGCTCAAGCATCCTCTGTT- CTGGCACAAAGGTCCTAGT   | 439 |
| L36156 | TCTTGCACTTGCT - - CAAGCATCCTCTGTT- CTGGCACAAAGGTCCTAGT   | 418 |
| X90692 | TCTTGCTCTTTCTGCTGAAATATCATCTGAT- CTGGCAAATGGTCCTACT<br>*****. *.*, **. *. * ****. . **** * ..**** ..       | 424 |
|        |  |     |
| L78163 | TGGCCAGTTCCATTAGGAAGAAGGGACAGCTTAACAGCAAACCGAACCCT   | 476 |
| U41657 | TGGCCAGTTCCATTAGGAAGAAGGGACAGCTTAACAGCAAACCGAACCCT   | 278 |
| X90693 | TGGAAGGTTCTTTTAGGAAGAAGAGATGGTTTTAACGGCAAACCGATTACT  | 487 |
| X90694 | TGGACGGTTCCTTTTAGGAAGAAGGGATGGTTTTAACCGCAAACCGAACACT   | 489 |
| L36156 | TGGACGGTTCCTTTTAGGAAGAAGGGATGGTTTTAACCGCAAACCGAACACT   | 468 |
| X90692 | TGGCAAGTTCCATTAGGAAGAAGGGATAGTTT GACAGCAAATAATTCCCT<br>*** .*****.*****.*** . * **.* ** ***** ... **       | 474 |
|        |  |     |
| L78163 | TGCAAATCAAAACCTTCCAGCACCTTTCTTCAA- - CCTCA- CTCAACTTA  | 523 |
| U41657 | TGCAAATCAAAACCTTCCAGCACCTTTCTTCAA- - CCTCA- CTCAACTTA  | 325 |
| X90693 | TGCTAATCAAAATCTTCCAGCTCC - - TTTCAATACTACTGATCAACTTA   | 534 |
| X90694 | TGCAAATCAAAATCTTCCGGCTCC - - ATTCAATTCCTTGGATCAACTTA   | 536 |
| L36156 | TGCAAATCAAAATCTTCCGGCTCC - - ATTCAATTCCTTGGATCACCTTA   | 515 |
| X90692 | TGCAGCTCAAAATCTTCCCTGCCCCACTTTCAA- - CCTTA- CTCGACTAA<br>***.. ***** *****.*** ** ***** * . . **. ***.     | 521 |
|        |  |     |
| L78163 | AAGCTTCCTTTG- CTGTTCAAGGTCTCAACACCCTTGATTTAGTTACACT  | 572 |
| U41657 | AAGCTTCCTTTG- CTGTTCAAGGTCTCAACACCCTTGATTTAGTTACACT  | 374 |
| X90693 | AAGCTGCATTTG- CTGCTCAAGGTCTCGATACTACTGATCTGGTTGCACT  | 583 |
| X90694 | AAGCTGCATTT- ACTGCTCAAGGCCTCAATACTACTGATCTAGTTGCACT  | 585 |
| L36156 | AA- CTGCATTTGACTGCTCAAGGCCTCATTACTCCTGTTCTAGTTGCCCT  | 564 |
| X90692 | AATCTAACTTTGA- TAATCAAAACCTCAGTACTACTGATCTAGTTGCACT<br>** *.*. *** *. *****.. ****.. ** **.* * *.*.*. * ** | 570 |
|        |  |     |
| L78163 | CTCAGGTGGTCATACGTTTGGAAGAGCTCGGTGCAGTACATTCATAAACC   | 622 |
| U41657 | CTCAGGTGGTCATACGTTTGGAAGAGCTCGGTGCAGTACATTCATAAACC   | 424 |
| X90693 | CTCCGGTGCTCATACATTTGGAAGAGCTCATTGCTCTTTATTTGTTAGCC   | 633 |
| X90694 | CTCGGGTGCTCATACATTTGGAAGAGCTCATTGCGCACAATTTGTTAGTC   | 635 |
| L36156 | CTCGGGTGCTCATACATTTGGAAGAGCTCATTGCGCACAATTTGTTAGTC   | 614 |
| X90692 | CTCAGGTGGCCATAACAATTGGAAGAGGTCAATGCAGATTTTTTCGTTGATC<br>*** ***** ..... ***** **..****. . **. *. .... *    | 620 |
|        |  |     |
| L78163 | GATTATACAACCTTCAGCAACACTGGAAACCCTGATCCAACCTCTGAACACA   | 672 |
| U41657 | GATTATACAACCTTCAGCAACACTGGA - - - CTGATCCA- CT- TGGACACA   | 468 |



|        |   |     |
|--------|---|-----|
| X90693 | GATTGTACAACTTCAGCGGTACGGGAAGTCCCGATCCAACCTCTTAACACA     | 683 |
| X90694 | GATTGTACAACTTCAGCAGTACTGGAAGTCCCGATCCAACCTCTTAACACA     | 685 |
| L36156 | GATTGTACAACTTCAGCAGTACTGGAAGTCCCGATCCAACCTCTTAACACA     | 664 |
| X90692 | GATTATACAATTTTCAGCAACACTGGAAACCCCGATTCAACTCTTAACACG     | 670 |
|        | *****.***** *****. . *.*** * *** ** * *..*****.         |     |
|        |   |     |
| L78163 | ACATACTTAGAAGTATTGCGTGCAAGATGCCCCCAGAATGCAACTGGGGA      | 722 |
| U41657 | ACATACTTAGAAGTATTGCGTGCAAGATGCCCCCAGAATGCAACTGGGGA      | 518 |
| X90693 | ACTTACTTACAACAATTGCGCACAATATGTCCCAATGGTGGACCTGGCAC      | 733 |
| X90694 | ACTTACTTACAACAAGTGGCGCACAATATGTCCCAATGGTGGACCTGGCAC     | 735 |
| L36156 | ACTTACTTACAACAAGTGGCGCACAATATGTCCCAATGGTGGACCTGGCAC     | 714 |
| X90692 | ACCTATTTACAACATTGCAAGCAATATGTCCCAATGGTGGACCTGGTAC       | 720 |
|        | ** ** * * * * * * * * . . . . . * * * * * * * * * * * . |     |
|        |   |     |
| L78163 | TAACCTCACCAATTTGGACCTGAGCACACCTGATCAATTTGACAACAGAT      | 772 |
| U41657 | TAACCTCACCAATTTGGACCTGAGCACACCTGATCAATTTGACAACAGAT      | 568 |
| X90693 | GAACCTTACCAATTTTCGATCCAACGACTCCTGATAAATTTGACAAGAACT     | 783 |
| X90694 | AAACCTTACCAATTTTCGATCCAACGACTCCTGATAAATTTGACAAGAACT     | 785 |
| L36156 | AAACCTTACCAATTTTCGATCCAACGACTCCTGATAAATTTGACAAGAACT     | 764 |
| X90692 | AAACCTAACCGATTTGGACCCAACACACCAGATACATTTGACTCCAAC        | 770 |
|        | .***** ****.***** * * * . * **.***.*** *****. * . *     |     |
|        |   |     |
| L78163 | ACTACTCCAATCTTCTGCAGCTCAATGGCTTACTTCAGAGTGACCAAGAA      | 822 |
| U41657 | ACTACTCCAATCTTCTGCAGCTCAATGGCTTACTTCAGAGTGACCAAGAA      | 618 |
| X90693 | ATTACTCTAATCTTCAAGTGAAAAAGGTTTGTCTTCAAAGTGATCAAGAG      | 833 |
| X90694 | ATTACTCCAATCTTCAAGTGAAAAAGGTTTGTCTTCAAAGTGATCAAGAG      | 835 |
| L36156 | ATTACTCCAATCTTCAAGTGAAAAAGGTTTGTCTTCAAAGTGATCAAGAG      | 814 |
| X90692 | ACTACTCCAATCTCCAAGTTGAAAGGGCTTGTCTTCAAGTGATCAAGAG       | 820 |
|        | * ***** ***** *.. . . *.** ** . * **.****** *****.      |     |
|        |   |     |
| L78163 | CTTTTCTCCACTCCTGGTGCTGATACCATTCCCATTTGTCAATAGCTTCAG     | 872 |
| U41657 | CGTTTCTCCACTCCTGGTGCTGATACCATTCC-ATTGTCAATAGCTTCAG      | 667 |
| X90693 | TTGTTCTCAACATCTGGTTCAGATACCATTAGCATTGTCAACAAATTCGC      | 883 |
| X90694 | TTGTTCTCAACTTCTGGTGCTGATACCATTAGCATTGTCAACAAATTCAG      | 885 |
| L36156 | TTGTTCTCAACTTCTGGTGCTGATACCATTAGCATTGTGACAAATTCAG       | 864 |
| X90692 | CTTTTCTCCAGAAATGGTTCGACACTATTTCTATTGTCAATAGTTTCGC       | 870 |
|        | . . ** * * * . *****.* ** * * * *****.* * . ***.        |     |
|        |   |     |
| L78163 | CAGTAACCAGAATACTTTCTTTTCCAACCTTTAGAGTTTCAATGATAAAAA     | 922 |
| U41657 | CG--AACCAGAATACTTTCTTTTCCAACCTTTAGAGTTTCAATGATAAAAA     | 715 |
| X90693 | AACCGATCAAAAAGCTTTTTTTGAGAGCTTTAGGGCTGCTATGATCAAAA      | 933 |
| X90694 | CACCGATCAAAAATGCTTTCTTTGAGAGCTTTAAGGCTGCAATGATTAAAA     | 935 |
| L36156 | CACCGATCAAAAATGCTTTCTTTGAGAGCTTTAAGGCTGCAATGATTAAAA     | 914 |
| X90692 | CAATAATCAAACCTCTCTTTGAAAATTTTGTAGCCTCAATGATAAAAA        | 920 |
|        | . . * * * . * * * * . * . * * * . . * .***** *****      |     |
|        |   |     |
| L78163 | TGGGTAATATTGGAGTGCTGACTGGGGATGAAGGAGAAATTCGCTTGCAA      | 972 |
| U41657 | TGGGTAATATTGGAGTGCTGACTGGGGATGAAGGAGAAATTCGCTTGCAA      | 765 |
| X90693 | TGGGAAATATTGGTGTGTTAACCAGGGAACCAAGGAGAGATTAGAAAACAA     | 983 |
| X90694 | TGGGCAATATTGGTGTGCTAACAGGGACAAAAGGAGAGATTAGAAAACAA      | 985 |
| L36156 | TGGGCAATATTGGTGTGCTAACAGGGACAAAAGGAGAGATTAGAAAACAA      | 964 |
| X90692 | TGGGTAATATTGGAGTTTTTAACCTGGATCTCAAGGTGAAATTAGAACACAG    | 970 |

|        |   |      |
|--------|---|------|
| L78163 | TGTAATTTTGTGAA---TGGAGACTCGT-----TTGGATTAGC   | 1007 |
| U41657 | TGTAATTTTGTGAA---TGGAGACTCGT-----TTGGATTAGC   | 800  |
| X90693 | TGCAACTTTGTTAATT-----CAA AATCAGCAGAACTTGGTCTTAT   | 1024 |
| X90694 | TGCAACTTTGTGAACTTTGTGAACTCAAATTCTGCAGAACTAGATTTAGC  | 1035 |
| L36156 | TGCAACTT-----TGTGAACTCAAATTCTGCAGAACTAGATTTAGC  | 1005 |
| X90692 | TG-----TAATGCTGTGAATGGGAATTCTTC-----TGGATTGGC   | 1005 |
|        | <div style="display: flex; justify-content: space-between; width: 100%;"> <span>**</span> <span>..</span> <span>. * . . *</span> </div> |      |

|        |   |      |
|--------|---|------|
| L78163 | TAGTGTGGCGTCCAAAGATGCTAAACAAAAGCTTGTGGCTCAATCTAAAT            | 1057 |
| U41657 | TAGTGTGGCGTCCAAAGATGCTAAACAAAAGCTTGTGGCTCAATCTAAAT            | 850  |
| X90693 | CAATGTTGCCTC - - - AGCAG - - ATTCATCTG - AGGAGGGTATGGTTAG - - | 1066 |
| X90694 | CACCATAGCATCCATAGTAG - - AATCATTAG - AGGATGGTATTGCTAGTG       | 1082 |
| L36156 | CACCATAGCATCCATAGTAG - - AATCATTAG - AGGATGGAATTGCTAGTG       | 1052 |
| X90692 | TACTGTAGTCACCAA - - - AG - - AATCATCAG - AAGATGGAATGGCTAGCT   | 1049 |
|        | *   *   *   *   *   *   *   *   *   *   *   *   *             |      |

|        |  |      |
|--------|--|------|
| L78163 | AAACCAATAATTAATGGGGATGTGCATGCTAGCTAGCATGTAAAGGCAAA | 1107 |
| U41657 | AAACCAATAATTAATGGGGATGTCGATGCTAGCTACGATGTAAAGGCAAA | 900  |
| X90693 | -----CTCAATGTAAA-TG-TAG                            | 1082 |
| X90694 | TAATATAAATAAATTAG-----CGTAAATGCACTTATTGAA-ATCTTG   | 1124 |
| L36156 | TAATATAAATAAATTAG-----CGAAAATGCACTTATTGAA-ATCTTG   | 1094 |
| X90692 | CATTCTAAAT--ATAAG-----CTTGGAATATTGAAGAGGTTCTAT     | 1090 |

. . . . \*

|        |  |      |
|--------|--|------|
| L78163 | TTAGGTTGTAAACCTCTTTGCTAGCTATATTGAAATAAACCAAAGGAGTA | 1157 |
| U41657 | TTAGGTTG-AAACCTCTTTGCTAGCTATATTGAAATAAACCAAAGGAGTA | 949  |
| X90693 | T--GATTGGAAGCAACTAA--TAAATTAAGAAGCTATAAC-----T     | 1119 |
| X90694 | T--GACTAGATGCCACTAA--TAAAT----AAGTTATAAC-----T     | 1157 |
| L36156 | T--GACTAGATCCCACTAA--TAAAT----AAGTTATAAC-----T     | 1127 |
| X90692 | A--ATTTTGTGCATACATA--TATGGTATGTG-----              | 1118 |

. . . \* . . . . \* . . . \*\* . . . . .

|        |  |      |
|--------|--|------|
| L78163 | GTGTGCATGTCAATTTCGATTTTGC-CATGTACCTCTTGGAATAT-----   | 1200 |
| U41657 | GTGTGCATGTCAATTTCGATTTTGC-CATGTACCTCTTGGAATATTATGTA  | 998  |
| X90693 | . ATGCACATT-CATGGTATGTGTGAGATAGTTATTAGATGCTTTGTGAGCA | 1168 |
| X90694 | AGGCACATTTTCATGTCACTTGAAATTTTCATGCCT-GTATATGAG-----  | 1200 |
| L36156 | AGGCACATTTTCATGTCACTTGAAATCCTATGCCTTGTATATTAGAGGACG  | 1177 |
| X90692 | ----CATGTGGTGTA--TTATGTTTTTGTATGTTCTTCAAGTTGATCA     | 1161 |
|        | * * . . . . . * . . . . . *                          |      |

|        |  |      |
|--------|--|------|
| L78163 | -----                                      | 1200 |
| U41657 | ATAATTATTTGAATCTC-----AAAAAAAAAAAAAAAAAAAA | 1031 |
| X90693 | AAAATCTTTTGGATTTC---ATTGAGTGTTTCT---       | 1200 |
| X90694 | -----                                      | 1200 |
| L36156 | TGT-TCTT-----C-----TTGGTATTATACTA--T       | 1200 |
| X90692 | GGGA-CTGTAGAAGCTCCCTAATAATATTTGTGTCAAAGT   | 1200 |



FIGURE 3B

|        |  |     |
|--------|--|-----|
| L78163 | MGSMRLLVVALLCAMHAGFSVSY---AQLTPTFYRETCPNLFPIVFGV     | 47  |
| U41657 | -----  | 0   |
| X90693 | MNSLRVAIALCCIV--VVLGGLPFSSNAQLDPSFYRNTCPNVSSIVREV    | 48  |
| X90694 | MNSL---ATSMWCVVLLVVLGGLPFSSDAQLSPTFYSKTCPTVSSIVSNV   | 47  |
| L36156 | M-----WCVLLVVLGGLPFSSDAQLSPTFYSKTCPTVSSIVSNV         | 40  |
| X90692 | MLGLSATA---FCCMVFLIGGVFSS-NAQLDPSFYNSTCSNLDSIVRGV    | 46  |
|        |  |     |
| L78163 | IFDASFTDPRIGASLMRLHFHDCFVQGCDSVLLNNTDTIESEQDALPNI    | 97  |
| U41657 | -----FHDCFVQGCDSVLLNNTDTIESEQDALPNI                  | 31  |
| X90693 | IRSVSKKDPRMLASLVRLHFHDCFVQGCDSVLLNKTDTVSEQDAFPNR     | 98  |
| X90694 | LTNVSKTDPRMLASLVRLHFHDCFVQGCDSVLLNNTATIVSEQQAFPNR    | 97  |
| L36156 | LTNVSKTDPRMLASLVRLHFHDCFVQGCDSVLLNNTATIVSEQQAFPNR    | 90  |
| X90692 | LTNVSQSDPRMLGSLIRLHFHDCFVQGCDSVLLNNTATIVSEQSAPPNR    | 96  |
|        | ***** ***.****.***.*** ** *                          |     |
|        |  |     |
| L78163 | NSIRGLDVVNDIKTAVENSCPDTVSCADILAIAAEIASVLGGGPGWPVPL   | 147 |
| U41657 | NSIRGLDVVNDIKTAVENSCPDTVSCADILAIAAEIASVAGRRSGWPVPL   | 81  |
| X90693 | NSLRGLDVVNQIKTAVEKACPNTVSCADILALSAELSSTLADGPDWKVPL   | 148 |
| X90694 | NSLRGLDVVNQIKLAVEVPCPNTVSCADILALAAQASSVLAQGPSWTVPL   | 147 |
| L36156 | NSLRGLDVVNQIKTAVESACPNTVSCADILALA-QASSVLAQGPSWTVPL   | 139 |
| X90692 | NSIRGLDVINQIKTAVENACPNTVSCADILALSAEISSDLANGPTWQVPL   | 146 |
|        | **.*****.*** ***.*****.*** ** *                      |     |
|        |  |     |
| L78163 | GRRDSL TANRTLANQNLPAPFFNL TQLKASFAVQGLNTLDLVTLSGGHTF | 197 |
| U41657 | GRRDSL TANRTLANQNLPAPFFNL TQLKASFAVQGLNTLDLVTLSGGHTS | 131 |
| X90693 | GRRDGLTANQLLANQNLPAPFNTTDQLKAAFAAQGLD TDLVALSGAHTF   | 198 |
| X90694 | GRRDGLTANRTLANQNLPAPFNSLDQLKAAFTAQGLNTD TDLVALSGAHTF | 197 |
| L36156 | GRRDGLTANRTLANQNLPAPFNSLDHLKLHLTAQGLITPVLVALSGAHTF   | 189 |
| X90692 | GRRDSL TANNSLAAQNLPAPTFNLTRLKSNFDNQNLS TDLVALSGGHTI  | 196 |
|        | ****.*****. **.******.*** ** * **.******             |     |
|        |  |     |
| L78163 | GRARCSTFINRLYNFSNTGNPDPTLNTTYLEVLRARCPQATGDNL TNLD   | 247 |
| U41657 | GRARCSTFINRLYNFSNTGLIH--LDTTYLEVLRARCPQATGDNL TNLD   | 179 |
| X90693 | GRAHCSLFVSRLYNFSGTGSPDPTLNTTYLQQLRTICPNGGPGTNLTNFD   | 248 |
| X90694 | GRAHCAQFVSRLYNFSSTGSPDPTLNTTYLQQLRTICPNGGPGTNLTNFD   | 247 |
| L36156 | GRAHCAQFVSRLYNFSSTGSPDPTLNTTYLQQLRTICPNGGPGTNLTNFD   | 239 |
| X90692 | GRGQCRFFVDRLYNFSNTGNPDSTLNTTYLQTLQAICPNGGPGTNLTDL    | 246 |
|        | ***.*** **.******.*** **.******.*** **.******        |     |
|        |  |     |
| L78163 | LSTPDQFDNRYYSNLLQLNGLLQSDQELFSTPGADTIPIVNSFSSNQNTF   | 297 |
| U41657 | LSTPDQFDNRYYSNLLQLNGLLQSDQERFSTPGADTIPLSIA-SANQNTF   | 228 |
| X90693 | PTTPDKFDKNYYSNLQVKKGLLQSDQELFSTSGSDTISIVNKFATDQKAF   | 298 |
| X90694 | PTTPDKFDKNYYSNLQVKKGLLQSDQELFSTSGADTISIVNKFSTDQNAF   | 297 |
| L36156 | PTTPDKFDKNYYSNLQVKKGLLQSDQELFSTSGADTISIVDKFSTDQNAF   | 289 |
| X90692 | PTTPDTFDSNYYSNLQVGKGLFQSDQELFSRNGSDTISIVNSFANNQTLF   | 296 |
|        | .*** **.******.***.***** ** *.*****.*** **.*         |     |
|        |  |     |
| L78163 | FSNFRVSMIKMGNIGVLTGDEGEIRLQCNFVN----GDSFGLASVAS-K    | 341 |
| U41657 | FSNFRVSMIKMGNIGVLTGDEGEIRLQCNFVN----GDSFGLASVAS-K    | 272 |
| X90693 | FESFRAAMIKMGNIGVLTGNQGEIRKQCNFVN---SKSAELGLINVAS-A   | 344 |
| X90694 | FESFKAAMIKMGNIGVLTGTKGEIRKQCNFVNFNVSNSAELDLATIASIV   | 347 |
| L36156 | FESFKAAMIKMGNIGVLTGTKGEIRKQCNFVN---SNSAELDLATIASIV   | 336 |

X90692 FENFVASMIMGNIGVLTGSQGEIRTQCNAV-----GNSSGLATVVT-K 340  
\*...\*...\*\*\*\*\*...\*\*\*\*\* \*\* ...\*....

|        |             |     |
|--------|-------------|-----|
| L78163 | DAKQKLVAQSK | 352 |
| U41657 | DAKQKLVAQSK | 283 |
| X90693 | DSSEEGMVSSM | 355 |
| X90694 | ESLEDGIASVI | 358 |
| L36156 | ESLEDGIASVI | 347 |
| X90692 | ESSEDGMASFF | 351 |
|        | ... ..      |     |

FIGURE 4



Diagram illustrating the DNA constructs used for the study. The top construct is a 17 kb BamHI-XbaI-BamHI construct. The bottom construct is a 3.3 kb SacI construct. The diagram shows the relative positions of the restriction sites and the size of the constructs. Dotted lines connect the BamHI sites to the sequence alignment below.

Sequence alignment showing the overlap between the two constructs:

```

OX312(ep)...TAAATCATAT ----- CAGCTTACTCC...
OX347(Ep)...TAAATCATATTAACT...//...ATGCTCAGCTTACTCC...
               |                       |
             -10                     +79
  
```

FIGURE 6

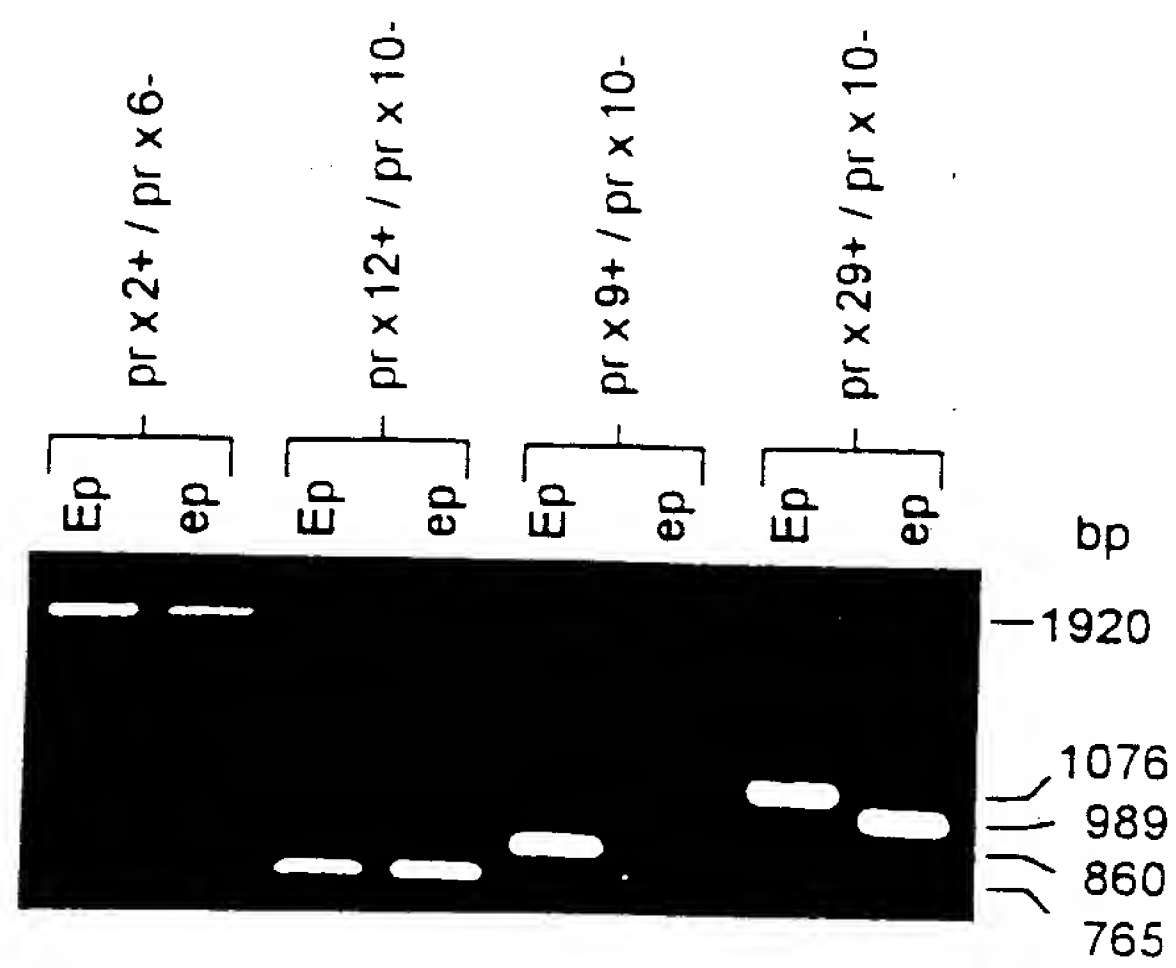




FIGURE 7

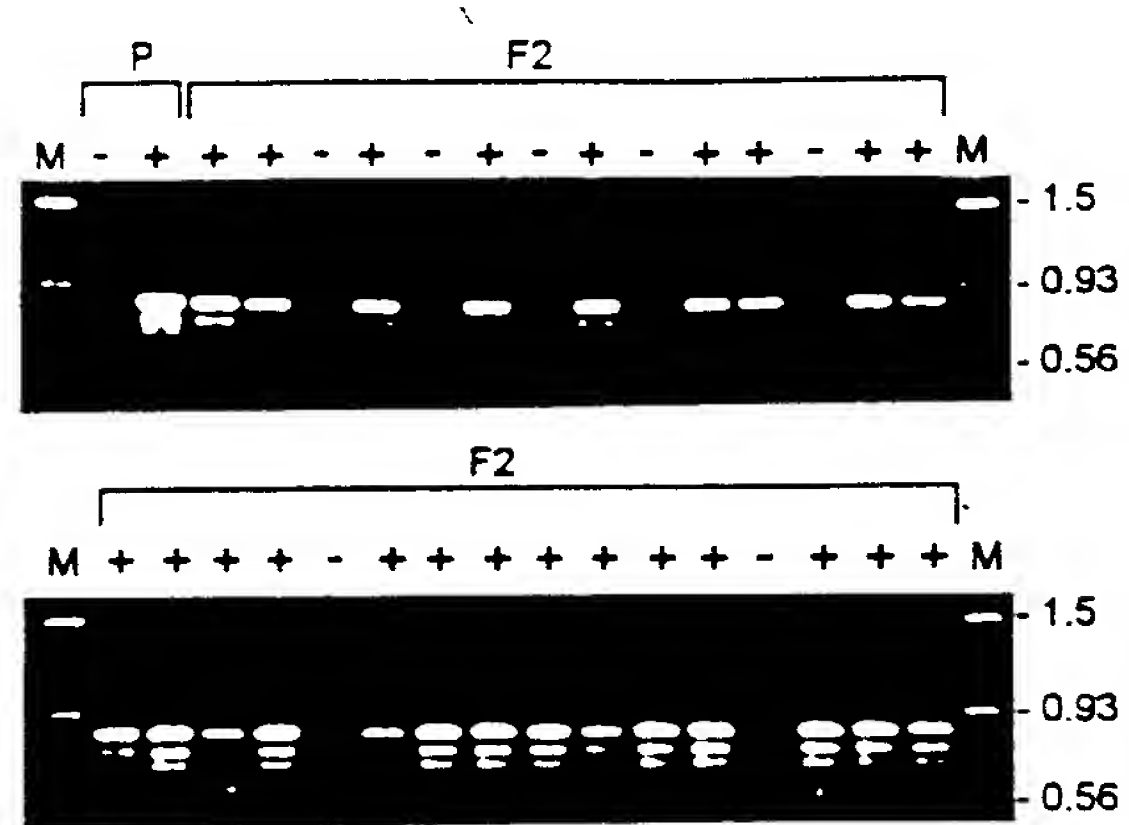


FIGURE 8

